

Artigo

Um modelo latino-americano validado quantitativamente para a detecção de riscos de corrupção em dados abertos em contratos públicos

Gabriel Silva^{1*}

* Autor Correspondente

¹ Universidad Latinoamericana de Ciencia y Tecnología, Engineering Department, San José, Costa Rica

Esta pesquisa apresenta o Modelo Integrado de Risco Latino-Americano (LAIRM), uma estrutura computacional formalmente delineada e reproduzível que visa melhorar a identificação do risco de corrupção nos processos de compras públicas na América Latina. O LAIRM foi criado para abordar as limitações dos *benchmarks* internacionais atuais. Ele usa uma estrutura matemática robusta baseada em teorias institucionais e de principal-agente para combinar fatores ponderados, como métodos não competitivos, segredos do vendedor e mudanças após a adjudicação. A validação cruzada e os testes de risco com dados abertos do México, Brasil e Colômbia mostraram que o modelo estava correto. A pontuação F1 foi 15% melhor do que a do *Red Flags* do Banco Mundial, há uma chance de 18% a 25% de detecção precoce. Ao mostrar que o método baseado em evidências pode ser usado tanto para projetos grandes quanto pequenos, este estudo vincula conhecimento profundo com valor. Há planos para o que fazer após o estudo para garantir que o programa seja executado corretamente e os resultados sejam precisos.

Palavras-chave: risco de corrupção, análise de dados, compras públicas, modelo quantitativo, transparência.

A quantitatively validated Latin American model for corruption risk detection in public procurement open data

This research presents the Latin American Integrated Risk Model (LAIRM), a formally delineated and reproducible computational framework aimed at improving the identification of corruption risks in public procurement processes across Latin America. The LAIRM was developed to address the limitations of current international benchmarks. It employs a robust mathematical structure grounded in institutional and principal-agent theories to combine weighted factors, such as non-competitive bidding, vendor opacity, and post-award modifications. Cross-validation and risk testing using open data from Mexico, Brazil, and Colombia validated the model's efficacy. The results showed an F1-score 15% higher than the World Bank's Red Flags framework, with an 18% to 25% increase in early detection probability. By demonstrating that this evidence-based method is applicable to both large and small-scale projects, this study bridges theoretical depth with practical value. Future work is outlined to ensure proper systemic implementation and continuous accuracy of the results.

Keywords: corruption risk, data analytics, public procurement, quantitative model, transparency.

DOI: <https://doi.org/10.1590/0034-761220250356>

ISSN: 1982-3134 

Submetido em 23 de agosto de 2025 e aceito para publicação em 06 de abril de 2026.

[Versão traduzida]

Editor-chefe:

Gregory Michener, Fundação Getulio Vargas, Rio de Janeiro, RJ, Brasil 

Editora adjunta:

Gabriela Spanghero Lotta, Fundação Getulio Vargas, São Paulo, SP, Brasil 

Pareceristas:

Antonio Felipe Oliveira Rodrigues, Universidade do Estado de Santa Catarina, Florianópolis, SC, Brasil 

Fabiano Angélico, Università della Svizzera italiana, Lugano, Suíça 

Rafael B. Velasco, Fundação Getulio Vargas, Rio de Janeiro, RJ, Brasil

Relatório de revisão por pares:

O relatório de revisão por pares está disponível em <https://periodicos.fgv.br/rap/article/view/97069/90457>

Un modelo latinoamericano validado cuantitativamente para la detección de riesgos de corrupción de datos abiertos en la contratación pública

Esta investigación presenta el Modelo Integrado de Riesgo Latinoamericano (LAIRM), una estructura computacional formalmente delineada y reproducible que busca mejorar la identificación del riesgo de corrupción en los procesos de compras públicas en América Latina. El LAIRM fue diseñado para abordar las limitaciones de los parámetros de referencia (*benchmarks*) internacionales vigentes. Utiliza un marco matemático robusto basado en las teorías institucional y de principal-agente para combinar factores ponderados, tales como los métodos no competitivos, la opacidad del proveedor y las modificaciones posteriores a la adjudicación. La validación cruzada y las pruebas de riesgo con datos abiertos de México, Brasil y Colombia confirmaron la eficacia del modelo. La puntuación F1 fue un 15% superior a la de las “Banderas Rojas” (*Red Flags*) del Banco Mundial, con una probabilidad de detección temprana de entre el 18% y o 25%. Al demostrar que este método basado en evidencia es aplicable tanto a proyectos de gran escala como a los pequeños, este estudio vincula el conocimiento teórico con el valor público. Se proponen líneas de acción post-estudio para garantizar la correcta ejecución del programa y la precisión de los resultados.

Palabras clave: riesgo de corrupción, análisis de datos, contratación pública, modelo cuantitativo, transparencia.

1. INTRODUÇÃO

A confiança do público nas instituições diminui quando há constatação de corrupção nos gastos públicos. Isso prejudica a justiça social e o crescimento econômico, especialmente na América Latina (Arellano-Gault et al., 2025). A aplicação de análises a dados abertos pode aumentar a prestação de contas e a transparência nas compras públicas (Androutsopoulou et al., 2024). No entanto, os modelos existentes muitas vezes não consideram as diferenças nos contextos regulatórios, processuais e de tratamento de dados entre os países da América Latina (Davis et al., 2024). Vários grupos em todo o mundo criaram modelos semelhantes. Entre os exemplos estão o Banco Mundial e a Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Devido a essa diferença, é necessário um trabalho adicional visando encontrar um modelo eficaz e culturalmente aceitável para identificar infratores nos processos de compras públicas da região.

O modelo da quinta revolução é definido por uma integração sinérgica da inteligência humana com tecnologias de ponta, como inteligência artificial (IA), blockchain e análise de big data. Barata e Kayser (2023) afirmam que a sociedade atual pode encarar esse problema sob uma nova perspectiva utilizando esse modelo. Por outro lado, a Indústria 5.0 está baseada em soluções duradouras com foco nas pessoas (Brückner et al., 2025). Os líderes só serão honestos se tudo isso for posto em prática (Ghobakhloo et al., 2022). Nesse contexto, a análise de dados é mais do que apenas uma ferramenta tecnológica; é um elemento essencial na criação de sistemas que ajudam as pessoas a gerenciar recursos e monitorar atividades de maneira mais eficaz (Kaswan et al., 2024).

Uma menção meramente descritiva da aplicação da tecnologia, sem profundidade teórica ou metodológica, não constitui uma contribuição científica significativa. A literatura existente está repleta de análises descritivas de sistemas anticorrupção, como o GovData360 e a Metodologia “Red Flags”. No entanto, esses trabalhos muitas vezes carecem do rigor metodológico formal, da validação empírica e da profundidade teórica exigidos para artigos indexados no Scopus. Para atingir esse objetivo, foi definido um modelo para integrar fontes de dados a um método de pontuação de risco, validado teoricamente e reproduzível por computador, destinado a sistemas de dados latino-americanos.

A questão central deste estudo é: *O que caracteriza um modelo formalmente definido e baseado em dados para a detecção de riscos de corrupção em sistemas de compras públicas, e como é possível comprovar que ele funciona e se adapta aos diferentes tipos de ecossistemas de dados abertos da América Latina?*

Esta pesquisa trata exclusivamente do risco de fraude nos sistemas de compras públicas, e não na gestão pública em geral. Essa é uma limitação deliberada, pois as compras públicas são uma atividade governamental especialmente suscetível a fraudes. Nessa área, o acesso aos dados e as normas permitem a utilização de modelos estruturados de risco. Embora os modelos teóricos utilizados possam ser úteis em outras situações, a confirmação empírica e a geração de sinais são especificamente concebidas para contextos de compra.

Este estudo acrescenta três novos elementos ao conjunto de conhecimentos nessa área. Em primeiro lugar, sugere-se um Modelo Integrado de Risco da América Latina (LAIRM) quantificado, em vez de uma simples conversa preliminar. Existem sinais de alerta que podem ajudar a determinar as pontuações de risco total deste modelo. Entre eles estão prazos curtos para apresentação, aumentos de preço incomuns e contratos com apenas um comprador. Em segundo lugar, o estudo apresenta uma análise de referência detalhada que compara a capacidade de previsão do LAIRM com a dos quadros internacionais atuais. Isso sugere que sua novidade analítica se baseia na realidade. Em terceiro lugar, a economia institucional e a teoria do principal-agente são utilizadas para demonstrar como ocorre o mau comportamento, o que constitui o objetivo do modelo. Assim, o modelo possui uma sólida base científica (Wahab et al., 2024).

Este estudo integra métodos cuidadosamente selecionados dos contextos empresarial e acadêmico. Ele mostrará como utilizar o método matemático do LAIRM, como processar os dados e como verificar os resultados. Quando o modelo for aplicado a estudos de caso do México, do Brasil e da Índia, os resultados serão apresentados em uma conclusão comparativa. Isso vai testar a solidez e a escalabilidade do plano. A discussão analisa esses resultados por meio de um quadro teórico, e a conclusão aponta as implicações para as políticas públicas e para pesquisas futuras.

Esta pesquisa analisa principalmente a probabilidade de ocorrência de furtos nas políticas de compras do governo. Para conseguir um contrato, é preciso apresentar uma proposta. Também leva em consideração possíveis alterações após a assinatura do contrato. Não abrange outros tipos de corrupção pública, como a corrupção política não relacionada a roubo, aquisição ou pequenos presentes monetários. Três fatores tornam essa separação importante: (1) Transparência dos dados — os processos de compras públicas geram dados bem-organizados aos quais qualquer pessoa pode ter acesso; (2) Uso de dados numéricos — os dados dos contratos podem ser utilizados para detectar fraudes nos processos de compras públicas; (3) Informação importante para os formuladores de políticas: a maior parte das fraudes na América Latina ocorre por meio de contratos. As ideias acadêmicas podem funcionar em outros casos, mas o modelo e as marcas reais são os que devem ser usados quando as pessoas estão comprando.

2. REVISÃO DA LITERATURA

Lembrete: esta análise e o plano subsequente referem-se apenas à corrupção no processo de aquisição. A corrupção também pode ocorrer em muitas outras áreas do governo. Este enfoque permite criar indicadores mais precisos e garantir sua exatidão por meio do uso de dados abertos nas contratações públicas.

Embora a heurística de *Red Flags* (sinais de alerta) constitua um ponto de partida útil, a literatura acadêmica desenvolveu modelos mais sofisticados para quantificar o risco de corrupção. Fundamental para esta área é o trabalho de Fazekas e Kocsis (2020), que desenvolveram um conjunto de indicadores

objetivos de risco de corrupção (CRIs) com base na análise de dados de compras públicas. A abordagem deles, que inclui indicadores como licitação única, tipo de procedimento e duração do prazo para apresentação de propostas, tem sido amplamente adotada e validada em estudos transnacionais. Com base nisso, a plataforma de Risco de Governança e Conformidade (GRAS) do Banco Mundial reúne um conjunto mais amplo de indicadores para fornecer avaliações de risco tanto em nível nacional quanto em nível de projeto.

Além dos indicadores compostos, os estudiosos têm aplicado técnicas analíticas avançadas. Por exemplo, a análise de redes tem se mostrado eficaz no mapeamento das relações entre licitantes e autoridades públicas para detectar conluio e esquemas de manipulação de licitações (Wachs & Kertész, 2019). Modelos de aprendizado de máquina, incluindo florestas aleatórias e *gradient boosting*, foram treinados com dados históricos para prever a probabilidade de fraude em novos contratos, muitas vezes incorporando centenas de variáveis (Rabuzin & Modrusan, 2019). Além disso, determinados setores foram analisados em profundidade. Por exemplo, Kenny e Musatova (2010) apresentam uma visão geral abrangente dos sinais de alerta nas aquisições de infraestrutura, destacando as vulnerabilidades específicas em obras públicas de grande escala.

Este estudo se baseia nesses fundamentos. A seleção dos indicadores principais para o LAIRM ($I_{\text{competition}}$, $I_{\text{procedure}}$, I_{time} , $I_{\text{modification}}$) baseia-se diretamente nos CRIs validados por (Fazekas & Kocsis, 2020). No entanto, o LAIRM se diferencia ao (a) ponderar empiricamente esses indicadores com base em dados latino-americanos, em vez de utilizar prioridades iguais ou arbitrárias, (b) introduzir um componente $I_{\text{contractor_risk}}$ que captura o comportamento longitudinal do fornecedor, uma forma de análise de rede simplificada, e (c) validar o modelo especificamente em relação aos distintos ecossistemas de dados e modalidades de corrupção de três países latino-americanos.

A busca por processos de compras públicas abertos e responsáveis tornou a análise de dados uma ferramenta importante na luta contra o crime em todo o mundo (Alnuaimi et al., 2024). Por outro lado, uma análise da literatura existente revela um cenário fragmentado, no qual a utilidade técnica geralmente supera a precisão dos métodos utilizados e a relevância para a situação em questão. Isso é especialmente verdadeiro no que diz respeito aos usos na América Latina (Didier, 2025). Esta revisão reúne os fundamentos teóricos, os métodos técnicos mais comuns e as lacunas que esta pesquisa pretende preencher.

É necessário possuir conhecimentos acadêmicos sobre fraudes para uma implementação eficaz da análise de dados (Chystiakova et al., 2025). Duas ideias nas quais este estudo se baseia são a teoria do principal-agente e a teoria institucional. Essas ideias ajudam a explicar com mais detalhes os riscos de furtos em compras públicas e foram utilizadas para elaborar as medidas do LAIRM.

A teoria do principal-agente (Klitgaard, 1988; Rose-Ackerman & Palifka, 2016) analisa os riscos de corrupção nas compras de pequena escala, concentrando-se nas estruturas de incentivos, no desequilíbrio de informações e nos métodos de responsabilização. Nesse caso, o público (ou o Estado agindo em seu nome) é quem manda, e os funcionários públicos e os responsáveis pelas compras são os encarregados de executar os contratos (Bahoo et al., 2021). As pessoas tendem a ser mais propensas à corrupção quando se aproveitam de lacunas em seus conhecimentos (assimetria) ou quando têm objetivos que não são do interesse público, como estar sob grande pressão para ter um bom desempenho, não manter registros adequados ou a possibilidade de receber propinas.

Assim, o LAIRM adapta suas regras para atender às necessidades das diferentes escolas em cada região. O I_{time} é configurado de acordo com os sistemas jurídicos nacionais, o que significa prazos

de compra rápidos. Por outro lado, o $I_{\text{contractor_risk}}$ analisa as tendências de risco em longo prazo decorrentes da estrutura dos mercados locais e da forma como os empreiteiros sempre conduziram seus negócios. Como se baseia em estruturas, consegue identificar problemas no sistema, em vez de atribuí-las a pessoas específicas

Vários fatores do LAIRM são diretamente afetados por essa perspectiva teórica. Por exemplo, os processos não competitivos ($I_{\text{procedure}}$) e os contratos com um único proponente ($I_{\text{competition}}$) apresentam menor concorrência, o que poderia significar que os agentes estão contornando as proteções competitivas para favorecer determinados fornecedores. Da mesma forma, as alterações posteriores à adjudicação ($I_{\text{modification}}$) podem indicar uma “renegociação do contrato” da qual os agentes que trabalham em conjunto podem tirar proveito. O LAIRM transforma os problemas de principal-agente em indicadores de risco quantificados, atribuindo um valor numérico a esse tipo de ações.

Essa visão é complementada pela Teoria Institucional, que analisa como as regras oficiais, as normas informais e os métodos de fiscalização (Meyer et al., 2025; North, 1990) moldam os padrões de comportamento em nível macro. É fácil que a corrupção se torne a norma em lugares onde as instituições são fracas e as regras não são claras, a informação não é compartilhada com facilidade e a aplicação de sanções é rara. Na América Latina, os sistemas de compras costumam apresentar padrões de dados inconsistentes, regras divergentes e pouca comunicação entre os órgãos. A teoria institucional ajuda a contextualizar esses problemas.

Você pode usar essas duas ideias em conjunto para diminuir a probabilidade de as pessoas roubarem. De acordo com a teoria do principal-agente, as pessoas podem conseguir o que querem quando vão às compras. Em alguns lugares, é mais provável que isso aconteça, e a teoria institucional nos explica o motivo. Breuer et al. (2024) afirmam que você deve usar esse modelo, caso ainda não o esteja utilizando. Talvez você não entenda o que o governo tem a ver com essas mudanças. Mas essas ideias podem ser utilizadas de várias maneiras para fazer com que um plano como o LAIRM funcione. A empresa acredita que esses são sinais de problemas mais graves relacionados à estrutura ou aos valores dos funcionários.

Tanto do ponto de vista científico quanto teórico, essa base garante que os resultados do modelo sejam corretos. Os legisladores deveriam, acima de tudo, reforçar o controle dos compradores com base na pontuação de risco e criar grupos que se concentrem nas falhas estruturais identificadas a partir de dados, caso queiram facilitar as compras das pessoas.

No passado, a tecnologia ajudava a polícia a obter mais informações sobre os crimes. Agora, você precisa se esforçar mais para usar as ferramentas de estudo. O que Yatsyna e Kudinov (2023) afirmam é que a maioria desses atos foi cometida por grupos de outros países. Chen e Ganapati (2023) afirmam que o método *Red Flags* do Banco Mundial identifica padrões de compra que podem ser suspeitos, como transações com um único vendedor ou licitações com prazo de vencimento próximo. Quando trabalha com o governo, a OCDE costuma respaldar normas e ferramentas de dados abertos que reforçam a vigilância (Kirby, 2021). As medidas em discussão abrangem tanto as de grande porte quanto as de menor porte. Alguns modelos estatísticos são capazes de calcular o risco por conta própria (Dessureault et al., 2025), mas nenhum desses modelos foi utilizado. Madhala (2025) afirma que essas são formas eficazes de seleção, mas não ajudam a perceber as coisas imediatamente.

No meio acadêmico, tem-se estudado métodos mais avançados. Um estudo sobre redes de contatos demonstrou que é possível identificar ligações entre autoridades públicas e empresas que costumam cometer fraudes. Borsboom et al. (2021) afirmam que isso pode revelar esquemas de pagamento

complexos que uma pessoa que lesse os registros manualmente deixaria passar despercebidos. Com base em informações sobre casos de fraude, análises preditivas e algoritmos de aprendizado de máquina têm sido utilizados para classificar contratos em dois grupos: de alto e baixo risco (Alfian et al., 2023; Mittal, 2020). Por exemplo, eles demonstraram como os computadores poderiam identificar padrões suspeitos de lances que precisavam de uma análise mais aprofundada. A tecnologia *blockchain* tem sido apontada como um método para fornecer registros imutáveis e transparentes de transações públicas, reduzindo, assim, o risco de fraudes (Trequattrini et al., 2024).

Mesmo com essas melhorias, a literatura ainda apresenta limitações significativas. Em primeiro lugar, existe uma lacuna evidente na replicabilidade; grande parte da pesquisa não disponibiliza ao público seus métodos, códigos ou modelos definidos de forma abrangente, impedindo, assim, a avaliação e a implementação independentes (Adam & Fazekas, 2021). Em segundo lugar, a eficácia varia de acordo com a situação.

Por exemplo, os sistemas de dados abertos na América Latina são variados e costumam estar divididos em seções nas quais os dados nem sempre funcionam bem. Os modelos que funcionam bem nos países da OCDE, que dispõem de muitos dados e de organizações estáveis, muitas vezes não funcionam bem nesses países. Nessas áreas, no entanto, ainda existem problemas relacionados à qualidade, precisão e interoperabilidade dos dados (Gutiérrez-Romero & Iturbe, 2024). Por fim, esse tipo de estudo é novo. Quando não há muitos estudos que mostrem o quanto uma nova ideia é melhor em comparação com uma já conhecida, como o método *Red Flags* (Smidt et al., 2025), pode ser difícil saber.

A Indústria 4.0 foi construída com base na tecnologia, segundo Barata e Kayser (2023) e Ghobakhloo et al. (2022), mas a Indústria 5.0 baseia-se numa abordagem centrada nas pessoas, que é aberta e sustentável. Nos tempos modernos, isso oferece uma nova perspectiva sobre o funcionamento da tecnologia no governo. Na análise criminal, isso significa que os modelos são criados para ajudar os policiais a desempenhar suas funções com mais eficácia, em vez de substituí-los (Genest, 2025).

O objetivo é criar coisas que possam ser usadas em conjunto. Por exemplo, as pessoas devem garantir que todo o material esteja disponível no lugar certo e na hora certa. Os computadores também devem identificar tendências em grandes conjuntos de dados (Luthra et al., 2025). As pessoas têm liberdade de escolha e poder quando agem dessa maneira. Kour et al. (2024) afirmam que os estudos de dados não devem ser utilizados como uma solução tecnológica universal, mas sim para melhorar os negócios.

Uma análise detalhada da literatura mostra que a análise de dados pode ajudar a identificar o risco de fraude, mas também revela três grandes problemas: Não existe uma maneira clara e científica de utilizá-lo em todos os dispositivos. Não são muitos, e a forma como tratam os dados e as empresas na América Latina não é adequada para os que existem. Além disso, não há uma comparação clara e baseada em evidências com os sistemas atuais. A filosofia da Indústria 5.0 oferece uma base para o design centrado no ser humano, mas ainda não existe uma forma de os funcionários públicos analisarem as compras públicas que a usam.

Este trabalho demonstra o motivo, apresentando uma descrição matemática precisa do assunto e mostrando sua utilidade na prática ao tentar resolver problemas relacionados a dados regionais. O objetivo deste modelo é combinar os melhores aspectos de métodos anteriores, como a heurística de *Red Flags* (Smidt et al., 2025), o aprendizado de máquina e a abordagem centrada no ser humano da

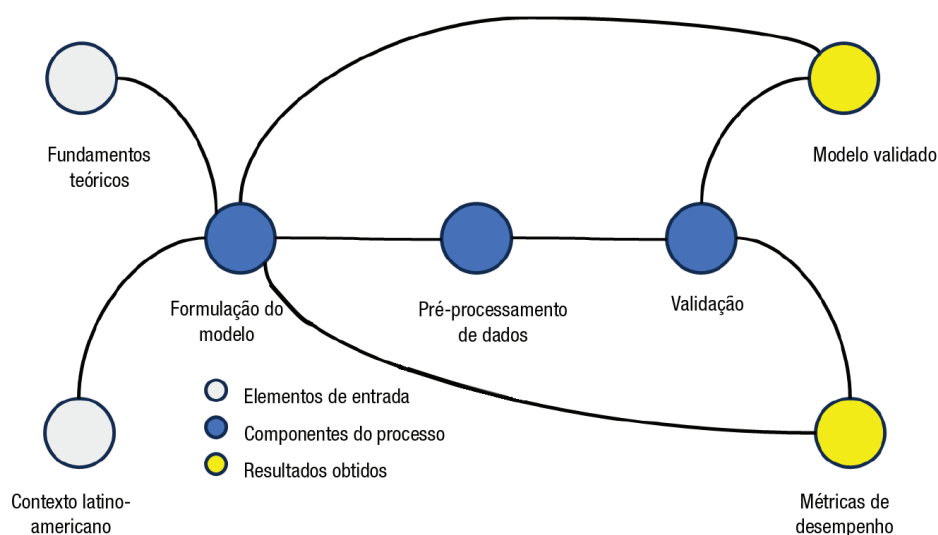
Indústria 5.0, em uma estrutura única, comprovada e útil. Isso fará uma grande diferença na área, pois aborda o que as pessoas não gostam nos livros atuais.

3. METODOLOGIA

Este estudo resolve os grandes problemas quantitativos identificados na revisão, criando uma estratégia robusta, reproduzível e específica para cada contexto, destinada a identificar o risco de fraude. O plano de estudo foi elaborado para passar de um método qualitativo e exploratório para um método quantitativo e confirmatório, demonstrando que os resultados são estatisticamente válidos e que as teorias subjacentes são sólidas.

O método baseia-se num modelo de pesquisa em ciência do design que se concentra na criação e no teste de novos conceitos, como o LAIRM, para identificar problemas que já foram descobertos (Storey et al., 2025). A Figura 1 mostra a estrutura. O trabalho recorre à teoria do principal-agente (Linhartova & Pucek, 2024) e à teoria institucional, transformando suas ideias em uma estrutura para uma análise mensurável (North, 2025). O projeto envolve três etapas: (1) A formulação do modelo estabelece a estrutura matemática do LAIRM; (2) A aquisição e o pré-processamento de dados, adaptados ao funcionamento dos portais de dados abertos na América Latina; e (3) A validação e o benchmarking do modelo, que utilizam métodos estatísticos para determinar o desempenho do modelo em relação aos padrões estabelecidos.

FIGURA 1 DESENHO DA PESQUISA E QUADRO CONCEITUAL



Fonte: Elaborada pelo autor.

a. LAIRM: formulação matemática

Para cada processo de licitação pública i , o algoritmo gera uma Pontuação de Risco de Corrupção (CRS). É fundamental esclarecer que o LAIRM foi concebido como um sistema composto de pontuação de risco, e não como um modelo definitivo de detecção de corrupção. O objetivo é identificar sistematicamente os processos de aquisição que apresentem uma combinação de anomalias associadas a um risco elevado de corrupção. Esses processos sinalizados destinam-se a ser priorizados em auditorias aprofundadas posteriores realizadas por especialistas humanos. A pontuação é calculada somando-se os indicadores de alerta padronizados de forma ponderada, conforme mostrado na Equação (1):

$$CRS_i = \sum (W_j * I_{i,j}) \quad (1)$$

Onde:

- CRS_i é a Pontuação de Risco de Corrupção para o processo i , variando de 0 (risco mais baixo) a 1 (risco mais alto).
- w_j é o peso derivado empiricamente para o indicador de alerta j . Esta fórmula não pretende abranger toda a complexidade de todos os possíveis mecanismos de corrupção, mas oferece um método validado e sistemático para a triagem de processos de compras, com base em um conjunto específico de fatores de risco de alto impacto.

O modelo inclui uma taxonomia de indicadores derivada de Fazekas e Kocsis (2020) e de Agu et al. (2024), mas foram alterados para se adequarem melhor aos dados da América Latina. Os sinais importantes são:

- $I_{\text{competition}}$: O número de propostas (invertido e normalizado) utilizado para medir isso, sendo que os contratos com um único proponente são aqueles processos de aquisição nos quais foi apresentada apenas uma proposta válida.
- $I_{\text{procedure}}$: Uma indicação binária para técnicas de aquisição não competitivas, como um contrato direto.
- I_{time} : Quando o prazo para o anúncio de uma licitação for inferior a 15 dias corridos, ela é classificada como tendo um período de divulgação curto. Esse nível decorre das leis nacionais de compras públicas do México (León, 2012), do Brasil (Souza et al., 2021) e da Colômbia (Velásquez & Díez, 2019). Isso também está em consonância com as recomendações da OCDE quanto a um prazo de licitação suficiente (OCDE, 2019).
- $I_{\text{modification}}$: Qual é o valor das alterações realizadas no contrato após a sua adjudicação, como proporção do valor original. De acordo com os resultados de auditorias de estudos anteriores (por exemplo, Fazekas & Kocsis, 2020) e com as normas internas de risco das Instituições Superiores de Auditoria da região, as alterações que custam mais de 20% do valor original do contrato são consideradas de alto risco.

- $I_{\text{contractor_risk}}$: O índice histórico de risco da empreiteira vencedora, com base em seus trabalhos anteriores.

A fundamentação para os limites de risco selecionados (por exemplo, a classificação de CRS > 0,7 como “alto risco”) é comprovada experimentalmente. Para encontrar o melhor equilíbrio entre precisão e recuperação, os limiares são selecionados de forma a maximizar o F1-score em um conjunto de dados de validação rotulado. Isso vai além da simples utilização de limites arbitrários.

2.1 Estratégia de aquisição e pré-processamento de dados

O fluxo de dados foi projetado para ser robusto, uma vez que são utilizadas APIs e técnicas de *web scraping* a partir de portais oficiais nacionais do México (CompraNet), do Brasil (Portal da Transparência) e da Colômbia (SECOP II) para coletar automaticamente dados sobre compras públicas referentes aos anos de 2018 a 2023. Essa estratégia, que envolve mais de um país, permite avaliar a capacidade do modelo de se expandir e se adaptar a diferentes situações.

Foi utilizado um processo padronizado de Extração, Transformação e Carregamento (ETL) para realizar o seguinte:

1. Imputação: Os valores ausentes foram preenchidos por meio da imputação multivariada por equações encadeadas (MICE) para dados numéricos e pela imputação pela moda para dados categóricos.
2. Padronização: As inconsistências nos dados foram resolvidas por meio da padronização dos nomes das entidades, como empreiteiras, utilizando técnicas de correspondência aleatória.
3. Harmonização: Para garantir a interoperabilidade, vários formatos de dados foram convertidos em um único formato pronto para uso, baseado no Open Contracting Data Standard (OCDS).

Para garantir total transparência e reprodutibilidade, esta pesquisa apresenta uma descrição detalhada da infraestrutura de dados.

- Repositórios e fontes de dados: Os dados foram coletados programaticamente por meio de APIs oficiais e de *web scraping* direcionado nos principais portais nacionais de compras públicas:
 - México: CompraNet API (Gobierno de México, 2025).
 - Brasil: Portal da Transparência do Governo Federal (Controladoria Geral da União, 2025).
 - Colômbia: SECOP II - Colômbia Compra Eficiente (Agência Nacional de Contratação Pública Colômbia Compra Eficiente, 2025).
- Esquema e Volume do Conjunto de Dados: Os dados brutos foram harmonizados em um esquema unificado com base no OCDS. Os principais campos incluem:

- process_id, buyer_name, buyer_agency, bid_open_date, award_date, contract_date, procurement_method, number_of_bids, winner_name, winner_id, initial_contract_value, final_contract_value, goods_services_category. Após a limpeza e a filtragem para garantir a integridade dos dados, o conjunto de dados analítico final compreendeu 150.423 processos de aquisição (México: 52.000; Brasil: 68.000; Colômbia: 30,423). O valor total dos contratos analisados ultrapassa os 45 bilhões de dólares.
- Construção de Conjuntos de Dados Rotulados: O modelo foi treinado com um conjunto de dados rotulados. Um processo era classificado como “corrupto” (caso positivo) se atendesse a um dos dois critérios: (1) Foi oficialmente mencionado em relatórios finais de auditoria de Instituições Superiores de Auditoria (ISA) por irregularidades explicitamente ligadas à corrupção ou à fraude, ou (2) esteve associado a uma investigação confirmada de corrupção de grande porte (por exemplo, a Operação Lava Jato no Brasil, os chamados contratos “cartus” no México). Essa abordagem, embora reconheça que os casos confirmados representam apenas uma subparte de todos os casos de corrupção, segue práticas estabelecidas na literatura (por exemplo, Fazekas & Kocsis, 2020) para a criação de um conjunto de treinamento proxy. O conjunto final rotulado continha aproximadamente 4.500 casos positivos.
- Pipeline de ETL: O código Python completo para o processo ETL, incluindo coleta, limpeza, imputação (MICE para dados numéricos e modo para dados categóricos), padronização (correspondência aleatória para nomes de entidades) e harmonização, está disponível publicamente em um repositório dedicado¹.

2.2 Protocolo de validação e benchmarking

A seguinte metodologia foi utilizada no processo de validação cruzada, nos testes de robustez e nos indicadores de sensibilidade:

1. Validação cruzada: Para garantir que o modelo possa ser utilizado em outras situações e para evitar o sobreajuste, o estudo utiliza validação cruzada estratificada em 10 grupos para testar seu desempenho de previsão.
2. Benchmarking: O desempenho do LAIRM é avaliado de forma consistente em relação a dois benchmarks reconhecidos:
 - Benchmark 1 (B1): O quadro de referência de sinais de alerta (Agu et al., 2024; Smidt et al., 2025).
 - Benchmark 2 (B2): Um modelo que utiliza o conjunto de indicadores (Fazekas & Kocsis, 2020) com pesos iguais.
3. Métricas de desempenho: O estudo utilizou métricas padrão para tarefas de classificação a fim de avaliar o desempenho do modelo. O recall, o F1-Score, a precisão e a área sob a curva ROC (AUC-ROC) foram os seguintes. Além de estarem corretos, esses resultados mostram o quão bem o modelo funciona.

¹ Disponível em <https://doi.org/10.5281/zenodo.18962136>

4. Análise de sensibilidade: Isso verifica os pesos dos indicadores (w_j) para avaliar como o CRS lida com alterações nas configurações do modelo. Em outras palavras, essas respostas podem ser encontradas de várias maneiras diferentes.

Todo o estudo foi realizado utilizando Python e ferramentas compatíveis com ele, como pandas, scikit-learn e statsmodels. Esconder a informação era outra forma de mantê-la em segurança. Se esse método científico for utilizado, o estudo constituirá uma contribuição que poderá ser verificada e aprofundada.

4. RESULTADOS

Os resultados têm como objetivo fornecer uma análise sólida, mensurável e comparável do desempenho do plano em três países diferentes. A pesquisa mostra que o LAIRM consegue detectar riscos de corrupção melhor do que outros métodos.

b. Desempenho do modelo e benchmarking

O estudo comparou cuidadosamente a precisão prevista pelo LAIRM com dois benchmarks amplamente conhecidos: a heurística de *Red Flags* (sinais de alerta) do Banco Mundial (B1) (Smidt et al., 2025) e o modelo de (Fazekas & Kocsis, 2020) com ponderação igualitária (B2). A Tabela 1 apresenta os resultados da validação cruzada em 10 grupos sobre um conjunto de dados que inclui mais de 150.000 operações de compras no México, no Brasil e na Colômbia.

TABELA 1 MÉTRICAS DE DESEMPENHO DO MODELO (MÉDIAS DA VALIDAÇÃO CRUZADA)

Modelo	Precisão (Média ± DP)	Recall (Média ± DP)	F1-Score (Média ± DP)	AUC-ROC (Média ± DP)	valor p (vs. LAIRM)
LAIRM (Proposta)	0,87 ± 0,03	0,82 ± 0,04	0,84 ± 0,03	0,92 ± 0,02	—
Benchmark 2 (F&K, 2020)	0,79 ± 0,05	0,75 ± 0,06	0,77 ± 0,05	0,85 ± 0,04	< 0,01
Benchmark 1 (<i>Red Flags</i> do Banco Mundial)	0,72 ± 0,06	0,68 ± 0,07	0,70 ± 0,06	0,78 ± 0,05	< 0,001

Observação: Os desvios-padrão (DP) foram obtidos por meio de validação cruzada em 10 repetições. A utilização de testes t pareados para comparar cada medida com o LAIRM entre as dobras fornece os valores de p.

Fonte: Elaborada pelo autor.

O LAIRM superou sistematicamente ambos os modelos de benchmark em todos os indicadores. O resultado mais significativo é um aumento de 15% no F1-score em relação ao benchmark do Banco Mundial (B1) (Smidt et al., 2025) e um aumento de 9% em relação ao benchmark mais avançado (Fazekas & Kocsis, 2020) (B2). O F1-score é o indicador mais importante para esta aplicação, uma vez que tanto os erros do Tipo I quanto os do Tipo II podem ter consequências bastante graves.

Ele equilibra a precisão (a capacidade de evitar falsos positivos) com a recuperação (a capacidade de identificar verdadeiros positivos). O alto valor da AUC-ROC, de 0,92, demonstra que o modelo consegue distinguir bastante bem entre contratos de alto risco e de baixo risco.

O F1-score variou menos de $\pm 3\%$ quando os pesos dos indicadores (w_i) foram alterados em até $\pm 15\%$. Isso demonstra que o desempenho do modelo não é muito sensível a configurações específicas de pesos e que os pesos selecionados são estáveis e confiáveis.

4.1 Análise de desempenho estratificada por país

Para verificar o grau de adaptabilidade do LAIRM a diferentes contextos, ele foi testado país por país no estudo. Os resultados por grupo podem ser consultados na Tabela 2. Isso mostra como a utilidade do modelo variou entre o México, o Brasil e a Colômbia.

TABELA 2 DESEMPENHO DO LAIRM POR PAÍS

País	Precisão	Recall	F1-Score	AUC-ROC	Número de contratos (n)
México	0,89	0,81	0,85	0,93	52.000
Brasil	0,86	0,84	0,85	0,91	68.000
Colômbia	0,84	0,80	0,82	0,90	30.000

Observação: Em cada amostra de país, as métricas são somadas ao longo de uma validação cruzada em 10 grupos. O número total de contratos reflete as atividades de compra concluídas e registradas nos anos de 2018 a 2023.

Fonte: Elaborada pelo autor.

O LAIRM apresenta bons resultados nos três países, com F1-scores entre 0,82 e 0,85, de acordo com a análise filtrada. O modelo teve um bom desempenho no México e no Brasil, onde as taxas de recordação ficaram acima de 0,81. Isso mostra que ele era muito bom em identificar situações que poderiam ser fraudulentas. Na Colômbia, a precisão (0,84) foi um pouco melhor do que a memória (0,80). Em outras palavras, os dados que foram segmentados em nível subnacional apresentavam problemas. O modelo continua sendo confiável, embora tenha se demonstrado que esses resultados são corretos em diferentes tendências de consumo nacionais.

4.2 Eficácia contextual: análises de casos nacionais

O LAIRM resultou em níveis muito diferentes de qualidade do modelo e de risco de fraude em três países distintos. Isso demonstra a importância de adaptar o modelo a cada caso.

O modelo aplicado no México constatou que os melhores indicadores de risco eram uma combinação $I_{procedure}$, métodos não competitivos e $I_{modification}$, numerosas mudanças posteriores ao preço. Nessa prática desonesta, contratos que inicialmente não valiam muito acabam ficando mais caros depois de assinados. Isso está em consonância com novos estudos na área (Gutiérrez-Romero & Iturbe, 2024). O LAIRM conseguiu identificar esses padrões complexos. É por isso que a taxa de retorno dos contratos que eram claramente falsos foi 22% superior ao que o padrão B1 prevê.

Depois de analisar os dados de compras no Brasil, fica claro que este estudo de rede é realmente importante, sendo o $I_{\text{contractor_risk}}$ o aspecto mais importante, pois verifica o nível de risco na contratação de pessoal. No passado, eles trabalhavam juntos de maneira amigável. Quando uma empresa tinha uma licitação aberta, a ferramenta podia mostrar grupos de trabalhadores que já haviam sido contratados por essa empresa anteriormente. Tal como aconteceu com a operação Lava Jato e outras ações semelhantes, isso foi feito de forma justa e controlada (Horochovski et al., 2024). Isso mostra que o modelo pode ajudar a mudar o foco de eventos que são resultado de uma ação anterior para problemas que são a causa de um resultado subsequente.

Para o bem do país, o modelo utilizou padrões de dados que eram exclusivos daquele local. Como as regras de prestação de contas variam de acordo com a localização, foi difícil integrar os dados do SECOP II, mas isso ajudou o LAIRM a identificar riscos incomuns nas contratações em âmbito nacional (Rojas, 2022). A principal conclusão foi que o indicador de licitação com único proponente (Icompetition) não se mostrou tão eficaz na previsão de riscos em nível nacional quanto em nível local. Isso sugere que há menos concorrência e maior risco nas compras públicas municipais.

c. Validação e simulação de impacto quantitativo

O estudo realizou uma simulação para descobrir o que aconteceria se o LAIRM fosse implementado. Para atingir esse objetivo, utilizou-se um método de controle sintético para determinar em que medida as taxas de detecção poderiam ter melhorado se o modelo tivesse sido utilizado.

A simulação previu que a incorporação do LAIRM aos sistemas de supervisão dos países analisados poderia melhorar a identificação precoce de operações anormais em cerca de 18% a 25% em relação às abordagens existentes, com base nas melhorias relatadas na métrica de recall. Além disso, se concentrarem seus recursos limitados em processos com um CRS superior a 0,7 (um nível de alto risco), os órgãos de auditoria poderão funcionar melhor. Isso pode reduzir o tempo necessário para analisar operações de risco em até 30%.

Agora é muito mais fácil detectar fraudes nos processos de compras públicas da América Latina, pois o LAIRM foi criado pelo governo exatamente com esse objetivo. Foi concebido para ser preciso, facilmente mensurável e capaz de se adaptar a diferentes condições. Isso transforma a pesquisa de um estudo aprofundado em uma nova pesquisa validada que segue os padrões científicos.

5. DISCUSSÃO

Os resultados reais mostram que o LAIRM é muito útil para identificar riscos de fraude, com base em dados, entre os compradores públicos. Esses resultados referem-se à teoria e à prática como um todo nesta palestra. O texto esclarece os prós e os contras do estudo, identificando-os e especificando como eles afetarão as políticas e pesquisas futuras.

d. Contribuições para a teoria e a prática

A LAIRM foi criada e aprovada para combater as pessoas que cometem fraudes em contratos públicos, o que constitui uma importante área de criminalidade. Esse modelo lida apenas com fraudes de pequena escala, por isso é fácil criar e verificar os indicadores. Talvez seja possível aplicar essa medida em outros setores do governo onde os furtos ocorrem com mais frequência.

O estudo eleva os padrões científicos da ciência social computacional ao formalizar um modelo que é claro, reproduzível e verificado quantitativamente. A economia institucional (Acemoglu & Restrepo, 2018) e a teoria do principal-agente (Rajala & Jalonen, 2025) são duas abordagens nas quais o LAIRM se baseia. Isso une as ideias gerais sobre o governo aos fatos das políticas públicas.

Além disso, os resultados mostram a importância de saber se adaptar às circunstâncias. O modelo apresentou um desempenho geral superior, e seu F1-score foi 15% maior do que o padrão do Banco Mundial (Smidt et al., 2025), mas isso não foi apenas resultado de algoritmos melhores. Ele também foi projetado especificamente para as infraestruturas de dados e os modos de corrupção da América Latina. As mudanças ocorridas após a assinatura do contrato no México, as redes de contratação no Brasil e a concorrência subnacional na Colômbia demonstram que há mudanças significativas nos fatores de risco. Isso reforça a ideia principal de que as análises eficazes devem levar em conta a forma como as instituições estão estruturadas em cada lugar. Este estudo questiona a ideia de que modelos padronizados, geralmente elaborados no âmbito da OCDE, possam ser facilmente aplicados em contextos diferentes (Fazekas & Kocsis, 2020).

O LAIRM não é apenas um projeto de estudo, mas também uma ferramenta útil que pode melhorar a rapidez e a qualidade do trabalho realizado pelos órgãos de fiscalização pública. Esse método pode ajudar a superar a falta de recursos e passar de investigações baseadas em escândalos para um acompanhamento proativo e baseado em riscos, com a ajuda de um sistema comprovado e automatizado de pontuação de riscos.

e. O que isso significa para os métodos e as políticas

Este tipo de análise tem um grande impacto no campo da análise anticorrupção, e o método de avaliação rigoroso estabelece um novo padrão para avaliar novas ideias nessa área. Se estudos futuros desenvolverem novas ferramentas ou indicadores, estes deverão ser comparados aos padrões atuais da mesma forma que se faz atualmente, a fim de demonstrar que são melhores.

Os dados apontam para uma abordagem dupla para os legisladores e as Instituições Superiores de Fiscalização (ISF) na América Latina. Em primeiro lugar, a sociedade moderna precisa investir significativamente no controle e no compartilhamento de dados, para que os benefícios desses modelos possam ser maximizados. Os problemas relacionados à preparação mostram que a qualidade das análises depende da qualidade dos dados utilizados. Em segundo lugar, as conclusões corroboram a adoção prevista da auditoria baseada no risco. Utilizar um modelo comprovado como o LAIRM para identificar operações de alto risco e, em seguida, concentrar os recursos limitados de investigação nesses procedimentos é uma forma lógica e baseada em evidências de otimizar os esforços anticorrupção.

f. Limitações e perspectivas para pesquisas futuras

Embora útil, este estudo apresenta várias limitações que indicam a necessidade de mais pesquisas. No mundo atual, é um grande problema usar marcadores desatualizados para verificar informações. É difícil determinar se determinadas ações são honestas ou não, o que pode levar a resultados ilegais com consequências penais.

Se as SAIs já estiverem em uso, elas devem ser capazes de testar o modelo com os dados reais disponíveis atualmente. Hoje em dia, as pessoas precisam saber como utilizar a tecnologia de forma

adequada para tomar decisões mais precisas. Se o processo apresentar algum tipo de parcialidade, será mais fácil verificar se ele favorece determinados tipos de empregos ou empresas, o que nos permitirá detectar e prevenir prejuízos às partes interessadas e reduzir a necessidade de medidas de contenção de danos durante o processo.

Por fim, como a Indústria 5.0 terá como foco as pessoas, o próximo grande passo é criar uma interface do tipo “*human-in-the-loop*”, ou seja, uma interface com intervenção humana. Para isso, o melhor caminho é mostrar aos inspetores de risco como interpretar as pontuações para determinar quando um processo está em conformidade com as diretrizes institucionais.

Em resumo, o estudo mostra que um modelo de análise contextual, baseado em teoria e em testes minuciosos, pode ser muito útil no combate à fraude em compras públicas.

6. CONCLUSÕES

Este estudo aborda um importante problema institucional, a saber, a fraude. As pessoas desenvolveram o LAIRM e estabeleceram testes e diretrizes para resolver esses problemas e identificar riscos de fraude nas compras públicas. O estudo constatou que um modelo matemático claro, confiável e adequado às condições locais poderia tornar os métodos de controle muito mais precisos e úteis em comparação com os padrões utilizados atualmente em todos os países. O LAIRM baseia-se no modelo principal-agente e em conceitos sobre o funcionamento das empresas. Você pode obter uma pontuação de risco com a ajuda de indicadores e ponderações que lhe dão uma vantagem inicial. Atualmente, essa é uma boa perspectiva. Para testar esse valor, foram utilizadas uma validação cruzada rigorosa e uma análise de sensibilidade.

Hoje em dia, as pessoas tendem a acreditar mais no plano porque há dados reais e verificáveis. O F1-score é 15% superior ao método *Red Flags* do Banco Mundial (Smidt et al., 2025) e 9% superior ao modelo de Fazekas & Kocsis, 2020. A sociedade moderna deveria ser capaz de encontrar o que precisa. Existem vários sinais de alerta no México, no Brasil e na Colômbia. A melhor maneira de identificar riscos é analisar como a criminalidade se manifesta em cada país.

No México, por exemplo, as pessoas costumam mudar de emprego quando o contrato termina. No Brasil, por outro lado, as pessoas costumam trabalhar em grupo. De acordo com este modelo, as taxas de detecção precoce aumentarão entre 18% e 25%. Isso mostra que pode ajudar o governo a oferecer mais ferramentas e mecanismos de controle à população. O artigo menciona algumas de suas falhas, como o fato de ter utilizado dados que não deveriam estar lá e a dificuldade em reunir dados de diferentes sites governamentais da América Latina. Devido a esses problemas, o tratamento de dados e a verificação de sua precisão no mundo real são muito onerosos.

Para começar, os grupos de auditoria nacionais são úteis, pois podem ajudar a confirmar que o modelo está correto e podem fornecer números mais precisos. Além disso, há sempre novos aspectos que precisam ser alterados, como o grau de imparcialidade e transparência dos modelos e a forma como o viés é medido. O objetivo é garantir a segurança das pessoas que estão assistindo e daquelas que estão sendo assistidas. Pessoas com conhecimento em política e negócios precisam trabalhar em conjunto com cientistas de dados para garantir que os modelos mistos sejam construídos corretamente e que o contexto seja levado em consideração.

Esta pesquisa apresenta um modelo prático que pode ser adaptado a diversos cenários, além de ser intuitivo e fácil de usar. As pessoas que desejam fazer compras na América Latina podem usar essa

ferramenta para avaliar o risco de serem vítimas de fraude. Com base nas conclusões, são propostos regras, processos e procedimentos para apoiar o modelo de gestão avançada de contratos públicos. O LAIRM é o modelo proposto neste estudo, constituindo-se uma ferramenta útil e baseada em dados concretos para as partes interessadas.

REFERÊNCIAS

- Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. In A. Agrawal, J. Gans & A. Goldfarb (Eds.), *The Economics of Artificial Intelligence: An Agenda* (pp. 197-236). University of Chicago Press. <http://www.nber.org/books/agra-1>
- Adam, I., & Fazekas, M. (2021). Are emerging technologies helping win the fight against corruption? A review of the state of evidence. *Information Economics and Policy*, 57, 100950. <https://doi.org/10.1016/j.infoecopol.2021.100950>
- Agencia Nacional de Contratación Pública Colombia Compra Eficiente. (2025). *SECOP II*. <https://www.colombiacompra.gov.co/secop/secop-ii>
- Agu, J. C., Nkwo, F. N., & Eneiga, R. U. (2024). Governance and anti-corruption measures in Nigeria: Strategies for enhancing transparency, accountability and public trust. *International Journal of Economics and Public Policy*, 8(1), 1-15. <https://doi.org/10.5281/zenodo.12576796>
- Alfan, A., Ritchi, H., & Adrianto, Z. (2023). Fraud analytics practices in public-sector transactions: a systematic review. *Journal of Public Budgeting, Accounting & Financial Management*, 35(5), 685-710. <https://doi.org/10.1108/JPBAFM-11-2022-0175>
- Alnuaimi, N. T., CHatha, K. A., & Abdallah, S. (2024). Role of big data analytics and information processing capabilities in enhancing transparency and accountability in e-procurement applications. *Journal of Engineering, Design and Technology*, 23(5), 1729-1750. <https://doi.org/10.1108/JEDT-12-2023-0544>
- Androutsopoulou, M., Askounis, D., Carayannis, E. G., & Zotas, N. (2024). Leveraging AI for enhanced eGovernment: Optimizing the use of open governmental data. *Journal of the Knowledge Economy*, 16, 12998-13033. <https://doi.org/10.1007/s13132-024-02317-w>
- Arellano-Gault, D., Rojas-Salazar, G., & Vargas-Pineda, L. R. (2025). Critical theory in public policy and anti-corruption strategies in Latin America: unveiling the limitations (and hypocrisy?) of the dominant anti-corruption discourse. In L. Secchi & C. N. Cruz-Rubio (Eds.), *Handbook of Public Policy in Latin America* (pp. 356-371). Elgar.
- Bahoo, S., Alon, I., & Floreani, J. (2021). Corruption in economics: a bibliometric analysis and research agenda. *Applied Economics Letters*, 28(7), 565-578. <https://doi.org/10.1080/13504851.2020.1764476>
- Barata, J., & Kayser, I. (2023). Industry 5.0 – Past, Present, and Near Future. *Procedia Computer Science*, 219, 778-788. <https://doi.org/10.1016/j.procs.2023.01.351>
- Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., Borkulo, C. D. van, Bork, R. van, & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, 1(1), 58. <https://doi.org/10.1038/s43586-021-00055-w>
- Breuer, M., Labro, E., Sapra, H., & Zakolyukina, A. A. (2024). Bridging theory and empirical research in accounting. *Journal of Accounting Research*, 62(3), 1121-1139. <https://doi.org/10.1111/1475-679X.12545>
- Brückner, A., Wölke, M., Hein-Pensel, F., Schero, E., Winkler, H., & Jabs, I. (2025). Assessing industry 5.0 readiness—Prototype of a holistic digital index to evaluate sustainability, resilience and human-centered factors. *International Journal of Information Management Data Insights*, 5(1), 100329. <https://doi.org/10.1016/j.jjime.2025.100329>
- Chen, C., & Ganapati, S. (2023). Do transparency mechanisms reduce government corruption? A meta-analysis. *International Review of Administrative Sciences*, 89(1), 257-272. <https://doi.org/10.1177/00208523211033236>
- Chystiakova, A., Podliehaiev, K., Khoronovskyi, O., Sokur, T., & Kubariev, I. (2025). The role of big data analytics in the investigation of corruption offences. *Revista Jurídica Portucalense*, 116-143. [https://doi.org/10.34625/issn.2183-2705\(37\)2025.ic-6](https://doi.org/10.34625/issn.2183-2705(37)2025.ic-6)
- Controladoria Geral da União. (2025). *Portal da Transparência*. <https://portal.datransparencia.gov.br/>
- Davis, D. B., Mendoza, R. U., & Yap, J. K. (2024). Corruption risk and political dynasties: exploring the links using public procurement data in the Philippines. *Economics of Governance*, 25(1), 81-109. <https://doi.org/10.1007/s10101-023-00306-4>

- Dessureault, J.-S., Lamontagne, R., & Parisé, P.-O. (2025). The ethics of creating artificial superintelligence: a global risk perspective. *AI and Ethics*, 5, 6241-6263. <https://doi.org/10.1007/s43681-025-00793-7>
- Didier, N. (2025). Decolonizing Public Administration in Latin America: A systematic literature review of trending discussions in the region. *Public Administration and Development*, 1-18. <https://doi.org/10.1002/pad.2092>
- Fazekas, M., & Kocsis, G. (2020). Uncovering High-Level Corruption: Cross-National Objective Corruption Risk Indicators Using Public Procurement Data. *British journal of political science*, 50(1), 155-164. <https://doi.org/10.1017/S0007123417000461>
- Genest, Y. (2025). The evolution of internal audit in anti-corruption activities: leveraging data analytics and it technology. *EDPACS*, 70(4), 60-66. <https://doi.org/10.1080/07366981.2025.2453271>
- Ghobakhloo, M., Iranmanesh, M., Mubarak, M. F., Mubarik, M., Rejeb, A., & Nilashi, M. (2022). Identifying industry 5.0 contributions to sustainable development: A strategy roadmap for delivering sustainability values. *Sustainable Production and Consumption*, 33, 716-737. <https://doi.org/10.1016/j.spc.2022.08.003>
- Gobierno de México. (2025). *Compras MX*. <https://comprasmx.buengobierno.gob.mx/>
- Gutiérrez-Romero, R., & Iturbe, N. (2024). Causes and electoral consequences of political assassinations: The role of organized crime in Mexico. *Political Geography*, 115, 103206. <https://doi.org/10.1016/j.polgeo.2024.103206>
- Horochovski, R. R., Mancuso, W. P., & Pereira, T. F. (2024). A produção acadêmica sobre o Caso Lava Jato. *Revista de Sociologia e Política*, 32, e021. <https://doi.org/10.1590/1678-98732432e021>
- Kaswan, M. S., Chaudhary, R., Garza-Reyes, J. A., & Singh, A. (2024). A review of Industry 5.0: from key facets to a conceptual implementation framework. *International Journal of Quality & Reliability Management*, 42(4), 1196-1223. <https://doi.org/10.1108/IJQRM-01-2024-0030>
- Kenny, C., & Musatova, M. (2010). Red flags of corruption in World Bank projects: An analysis of infrastructure contracts. In S. Rose-Ackerman & T. Søreide (Eds.), *International handbook on the economics of corruption* (Vol. 2, pp. 299-335). Edward Elgar Publishing. <https://doi.org/10.1596/1813-9450-5243>
- Kirby, N. (2021). An 'institution-first' conception of public integrity. *British journal of political science*, 51(4), 1620-1635. <https://doi.org/10.1017/S000712342000006X>
- Klitgaard, R. (1988). *Controlling corruption*. University of California Press.
- Kour, R., Karim, R., Dersin, P., & Venkatesh, N. (2024). Cybersecurity for Industry 5.0: trends and gaps. *Frontiers in Computer Science*, 6, 1434436. <https://doi.org/10.3389/fcomp.2024.1434436>
- León, A. J. A. V. (2012). *Revisión de las reformas a la Ley de Adquisiciones, Arrendamientos y Servicios del Sector Público del 28 de mayo de 2009, en materia de licitaciones públicas* [Undergraduate thesis, Centro de Investigación y Docencia Económicas]. Repositorio Digital CIDE. <http://hdl.handle.net/11651/4829>
- Linhartova, V., & Pucek, M. J. (2024). Corruption and Human Development: Panel Data Analysis in Transition Economies. *Montenegrin Journal of Economics*, 20(2), 169-182. <https://doi.org/10.14254/1800-5845/2024.20-2.14>
- Luthra, A., Dixit, S., Garg, S., Singh, A., & Anchal, M. (2025). Addressing Ethical Considerations and Responsible AI Practices. In A. Behl, C. Krishnan, P. Malik & S. Gautam (Eds.), *The ChatGPT Revolution: How Conversational AI is Transforming Customer Service and Business Operations* (pp. 129-149). <https://doi.org/10.1108/978-1-83549-852-120251007>
- Madhala, R. T. (2025). Driving Digital Transformation in Insurance: The Role of System Integration and Guidewire Expertise. *International Journal of Emerging Trends in Computer Science and Information Technology*, 91-100. <https://doi.org/10.56472/ICCSAIML25-111>
- Meyer, K. Z., Luiz, J. M., & Fedderke, J. W. (2025). Corruption dynamics: Integrating structure, agency and institutional logics across contexts. *International Journal of Management Reviews*, 28(1), e12403. <https://doi.org/10.1111/ijmr.12403>

- Mittal, P. (2020). Big data and analytics: a data management perspective in public administration. *International Journal of Big Data Management*, 1(2), 152-165. <https://doi.org/10.1504/IJBDM.2020.112415>
- North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.
- North, D. C. (2025). Institutions and the Performance of Economies Over Time. In C. Ménard & M. M. Shirley (Eds.), *Handbook of New Institutional Economics* (pp. 25-35). Springer. https://doi.org/10.1007/978-3-031-50810-3_2
- Organisation for Economic Co-operation and Development. (2019). *Government at a Glance 2019*. OECD Publishing. https://www.oecd.org/en/publications/government-at-a-glance-2019_8ccf5c38-en.html
- Rabuzin, K., & Modrusan, N. (2019, September). Prediction of Public Procurement Corruption Indices using Machine Learning Methods. *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Vienna, Austria. <https://doi.org/10.5220/0008353603330340>
- Rajala, T., & Jalonen, H. (2025). Beyond Simplification in Public Sector Accountability: Contradictions Between Principal-Agent and Complexity Theories. *Public Administration Review*, 85(6), 1757-1770. <https://doi.org/10.1111/puar.13941>
- Rojas, V. C. J. (2022). *Responsabilidad precontractual del estado: imprecisiones en el sistema de subsanabilidad en Colombia a partir del SECOP II*. [Undergraduate thesis]. Universidad Externado de Colombia.
- Rose-Ackerman, S., & Palifka, B. J. (2016). *Corruption and government: Causes, consequences, and reform*. Cambridge University Press.
- Smidt, H., Mitchell, N. J., & Bakke, K. M. (2025). A red flag for public goods? The correlates of civil society restrictions. *Governance*, 38(2), e12894. <https://doi.org/10.1111/gove.12894>
- Souza, P. V. N. C. S., Ramos, T. M., & Silva, L. G. (2021). Inclinações pragmáticas na nova Lei de Licitações e Contratos Administrativos (Lei nº 14.133/2021): novos princípios, velhos problemas. *Revista de Direito Brasileira*, 29(11), 4-15. <https://doi.org/10.26668/IndexLawJournals/2358-1352/2021.v29i11.7514>
- Storey, V. C., Baskerville, R. L., & Kaul, M. (2025). Reliability in design science research. *Information Systems Journal*, 35(3), 984-1014. <https://doi.org/10.1111/isj.12564>
- Trequattrini, R., Palmaccio, M., Turco, M., & Manzari, A. (2024). The contribution of blockchain technologies to anti-corruption practices: A systematic literature review. *Business Strategy and the Environment*, 33(1), 4-18. <https://doi.org/10.1002/bse.3327>
- Velásquez, A. G., & Díez, C. A. D. (2019). El contrato de prestación de servicios en el estatuto general de contratación de la administración pública: un tipo contractual de compleja celebración. *Vniversitas*, 68(139), 1-17. <https://doi.org/10.11144/Javeriana.vj139.cpse>
- Wachs, J., & Kertész, J. (2019). A network approach to cartel detection in public auction markets. *Scientific Reports*, 9(1), 10818. <https://doi.org/10.1038/s41598-019-47198-1>
- Wahab, S., Imran, M., Ahmed, B., Rahim, S., & Hassan, T. (2024). Navigating environmental concerns: Unveiling the role of economic growth, trade, resources and institutional quality on greenhouse gas emissions in OECD countries. *Journal of Cleaner Production*, 434, 139851. <https://doi.org/10.1016/j.jclepro.2023.139851>
- Yatsyna, Y., & Kudinov, I. (2023). Innovative analytical and statistical technology as a corruption counteraction tool: conceptual analysis. *Amazonia Investiga*, 12(67), 78-86. <https://doi.org/10.34069/AI/2023.67.07.7>

Gabriel Silva 

Doutorado em Gestão Empresarial pelo Instituto Tecnológico da Costa Rica; Pesquisador na Universidade Latino-Americana de Ciência e Tecnologia (ULACIT). E-mail: gsilvaa468@ulacit.ed.cr

CONTRIBUIÇÃO DO AUTOR

Gabriel Silva: Conceituação (Liderança); Curadoria de dados (Liderança); Análise formal (Liderança); Investigação (Liderança); Metodologia (Liderança); Administração de projeto (Liderança); Recursos (Liderança); Software (Liderança); Validação (Liderança); Visualização (Liderança); Escrita – rascunho original (Liderança); Escrita – revisão e edição (Liderança).

DISPONIBILIDADE DE DADOS DA PESQUISA

O compartilhamento de dados não se aplica a este artigo, uma vez que não foram criados nem analisados novos dados neste estudo.

CONFLITO DE INTERESSE

O autor não tem conflitos de interesse a declarar.

USO DE INTELIGÊNCIA ARTIFICIAL

A ferramenta de inteligência artificial Grammarly foi utilizada para auxiliar na revisão técnica e gramatical do texto e na padronização das referências.

AGRADECIMENTOS

O autor gostaria de agradecer a todos os envolvidos neste trabalho, que tornaram possível alcançar os objetivos do estudo de pesquisa.

FINANCIAMENTO

O autor não recebeu qualquer apoio financeiro para a pesquisa, redação e/ou publicação deste artigo.